Recent Results in Speech Recognition for the Tatar Language

Aidar Khusainov(✉)

Institute of Applied Semiotics of the Tatarstan Academy of Sciences,
Kazan Federal University, Kazan, Russia
khusainov.aidar@gmail.com

Abstract. This paper presents a comparative study of several different systems for speech recognition for the Tatar language, including systems for very large and unlimited vocabularies. All the compared systems use a corpus based approach, so recent results in speech and text corpora creation are also shown. The recognition systems differ in acoustic modelling algorithms, basic acoustic units, and language modelling techniques. The DNN based system with the sub-word based language model shows the best recognition result obtained on the test part of speech corpus.

Keywords: Speech recognition · Acoustic modelling · Language modelling · Tatar language

1 Introduction

The conventional way of building speech recognition systems is to obtain required acoustic models, a pronunciation dictionary, a language model, and use some of the decoders. The situation can be worse whenever you have to recognize the speech of an under-resourced language. In that case some (or all) of the required resources and algorithms may not exist.

In this article we present our recent results in creating continuous speech recognition systems for the Tatar language. We used the word based approach to create a very large vocabulary speech recognition system and the sub-word based approach for the case of unlimited vocabulary recognition.

Tatar is spoken by several million people and is the second spoken language in Russia. There are 4.2 million of speakers in Russia and near 5.2 million of speakers in the world [8]. The Cyrillic alphabet (unified in 1939) consists of 39 characters. There are 12 vowel and 28 consonant sounds. Different dialects of Tatar can be identified: Western, Kazan (Middle) and Eastern. Based on the existing language classification [7], in 2013 it was assigned to the under-resourced language class [6]. However, recent results in machine translation [1], speech analysis and synthesis [3] fields can change this situation. The main feature of Tatar in case of creating speech recognition systems is the agglutinative nature of the language, so one word can have tens of surface forms. It leads to a very high OOV rate problem.

K. Ekštein and V. Matoušek (Eds.): TSD 2017, LNAI 10415, pp. 183–191, 2017.
DOI: 10.1007/978-3-319-64206-2_21